2/25/80

LEVEL II

AD A096582

DTIC
ELECTE
MAR 20 1981
S D
F

Planning Systems Incorporated

7900 Westpark Drive • Suite 507

McLean, Virginia 22101

(703) 790-5950

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

80 9 3 080

DDC FILE COPY

LEVEL II

1

6 ON THE ESTIMATION OF SHIPPING DENSITIES FROM OBSERVED DATA,

10 J.Z./YAO
A.E./BARNES

Prepared for:

Office of Naval Research
Code 102-OSC:LRAPP
Arlington, Virginia 22217

Contract No. 15 N00014-73-C-0223

14 PSI-TR-004018

11 Apr[redacted]975

Planning Systems Incorporated
7900 Westpark Drive
Suite 507
McLean, Virginia 22101

390451

A

JCR

TABLE OF CONTENTS

I. Problem Definition

II. Weighted Average Algorithm

III. Algorithm for Estimating the Monthly Shipping Density Averages from a Partial Data Set

IV. Theory of Estimating Ship Density Time Averages from Incomplete Data

V. Applications to a Specific Case

VI. Theoretical Sensitivity Analysis

B

I. Problem Definition

A. Background

Passive surveillance system performance in terms of area coverage is ultimately limited by the noise environment in which it must operate. As the state of the art in design has progressed, interest is no longer exclusively in the omni-level of the noise field but now includes as equally important the spatial, temporal and directional characteristics of this field. Additionally, extreme concern with the low frequency portion of the noise spectrum develops as the need for enhanced classification capability increases. The predominant component of the low frequency noise field for most platforms is that of ship noise. Thus, as the need for increased passive detection and classification capability increases, one is forced to consider the low frequency portion of the spectrum. In turn, this implies the need to quantitatively specify the ambient noise environment.

Shipping distributions play a dominant role in the calculation of low frequency ambient noise, since these distributions affect both the level and the directional characteristics of this noise field. The geographical distribution of ships throughout the ocean basin gives rise to the directionality, both horizontal and vertical. The variation of transmission loss with range and bearing serves to accentuate or moderate this directional character, and also determines the level of the shipping contribution to ambient noise field at the receiver. Both the ambient noise and the transmission loss are a function of the depth of the receiver as substantiated by the comparison of simulation models and experimental results.

It is clear that all ships throughout an ocean basin contribute to the ambient noise field. These individual contributions will be more or less significant to the level and directionality of the ambient noise field depending upon the transmission loss from the ship to the receiver. Near ships, i.e., those within 50 miles, may not be more important than far ships, i.e., 200-500 miles, depending on properties of the various convergence zones. Calculations clearly show that both the level and directional characteristics of ambient noise are sensitive to the total shipping distribution, including both near and far ships.

Historical shipping fields are available and serve as a basis for planning and preliminary calculations in the determination of real time shipping distributions. Experience backed by empirical results demonstrates that for model validation and ambient noise determination for particular sites in a particular experiment an accurate and timely observation of the true total shipping distribution must be employed to obtain satisfactory results. In the near term this distribution is obtainable only by the use of real time aircraft surveillance.

This report addresses the problems and approaches involved in reducing this observational data obtained from aircraft surveillance in order to obtain a representative shipping field for the time frame of the experiment.

B. Basic Problems

The raw data, of course, is highly dependent upon the flight tracks and equipment performance. This implies that if the region of interest is divided into rectangles of some preset size (e.g., 1° x 1°) then the coverage afforded to these squares is somewhat uneven. Specifically,

on a flight day a square may have multiple or only partial coverage, i.e., a plane may cover the square more than once (by covering it on two legs of the flight), or may only cover a fraction of the square (if it is not directly under the flight path). Different flight days will usually utilize different tracks, resulting in differences in the area being covered. Various portions of the overall survey area have much higher priority than others (viz, a line containing the source and receiver ships in an acoustic experiment). Also, different planes will often be used. The equipment on board such planes can vary greatly in quality and reliability. The planes may even have different types of equipment. And different radar operators will account for more differences in data; a good operator can 'tweek' his set with fine adjustments to a high level of quality; another operator might not be able to do any 'fine tuning'.

Thus, the raw data gathered is not completely uniform; indeed in some cases quality variations may be extreme. Also there are historical shipping fields of various quality which may be utilized. From such an accumulation of data one wishes to construct, using an algorithm, a representative shipping field.

C. Techniques

There are innumerable possible ways of answering the question "How should one estimate density?" In light of the types of data available, the problems mentioned above, and the constraints on the analysis (viz, a limited number of surveillance flights and a rapid response in terms of the overall shipping field), there are two algorithms which

appear most suitable. The choice of which of these algorithms to use in analyzing a specific set of data depends on the attributes of that data set (e.g., the quality and quantity of the data, the conditions under which various subsets of it were taken). These algorithms are based on both theory and practice, and have been utilized in the analysis of surveillance data from exercises.

The first one is the weighted average algorithm, which is described in Section II. If the data is very limited, this algorithm should be used, which will provide the (weighted) mean and variance of the shipping density in each square. This algorithm is quite rapid on a programmable desk top electronic calculator. If more data is available, or the data is of very uneven quality, then the other algorithm should be utilized. This algorithm, which can utilize incomplete data, is described in Section III. The theory underlying this algorithm is presented in Section IV.

Like any mathematical or statistical tool, it can be misapplied. If data is uniformly poor or worthless, no algorithm will convert it into an accurate prediction (the computer maxim of GIGO is completely applicable). If data is very scanty (e.g., one day of observation), there is no point in utilizing such an algorithm; there is little or nothing to cross-correlate. In such a situation the former algorithm should be utilized.

Typical results of the weighted average algorithm are given in Section II. An example of the other algorithm as applied to actual data is given in Section V.

The description of the weighted average algorithm (Section II) is taken from PSI report TR-004002, while the description of the partial data set algorithm comes from TR-004012.

II. Weighted Average Algorithm

It is necessary to assure that the shipping field density during the course of the exercise is stationary, i.e., not time dependent. It is then of interest to consider the average density. A moments reflection will indicate that the arithmetic mean

$$A = \frac{1}{n} \sum_{1}^{n} d_i \qquad (1)$$

has certain deficiencies. For example, suppose on day 1 an area was covered at three different times by different aircraft, but on day 2 only one aircraft was in the area, and it only covered half the square. The above formula would attack equal significance to the computed values of the ship density on those two days, whereas the estimate of the density on day 1 is (statistically) superior to that of day 2. For this reason a weighted average will be used.

Before plunging into the relevant equations, several quantities will be defined. Fix an area of the ocean (for the computations 5° x 5° squares will be used), and let n_i be the number of ships observed on the ith day ($i = 1,2,\ldots,n$), and let p_i be the proportion of the area covered on the ith day. If $p_i \neq 0$ then an estimate of the density of ships in the area on the ith day is given by

$$d_i = n_i/p_i \qquad (2)$$

For later convenience, define

$$s = \sum_{i=1}^{n} p_i \qquad (3a)$$

$$s_2 = \sum_{i=1}^{n} p_i^2 \qquad (3b)$$

The weighted average has the form

$$d = \frac{\sum w_i d_i}{k} \tag{4}$$

where w_i and k are constants (the constant k is introduced so that the w_i may have simpler form). In order to attach the same importance to each square mile which was covered, the weight must be proportional to the amount of area covered, i.e., $w_i = p_i$. The constant k is chosen such that d is an unbiased estimator. That is, if d_i are independent samples of a random variable (called "density") having mean μ, then k is chosen such that the expected value of d is μ, i.e.,

$$E[d] = \mu \tag{5}$$

which implies $k = s$. But $w_i d_i = p_i d_i = n_i$ so

$$d = (\sum n_i)/s \tag{6}$$

For similar reasons, a weighted estimator of variance is desirable. This estimator has the form

$$\hat{\sigma}^2 = \frac{1}{\hat{k}} \sum \hat{w}_i (d_i - d)^2 \tag{7}$$

Letting $\hat{w}_i = p_i$ gives a system consistent with equation (6). This system favors the larger area coverage, but does not place inordinate emphasis on the days with maximum coverage. Choosing $\hat{k}$ such that $\hat{\sigma}^2$ is an unbiased estimator gives

$$\hat{\sigma}^2 = \frac{s}{s^2 - s_2} \left(\sum_{i=1}^{n} n_i^2/p_i - d^2 s \right) \tag{8}$$

where $n_i^2/p_i \equiv o$ when $n_i = p_i = o$. The (weighted) average density, by 5° x 5° square, is given in Figure 1, and the variance is given in Figure 2, for data gathered recently in the Pacific.

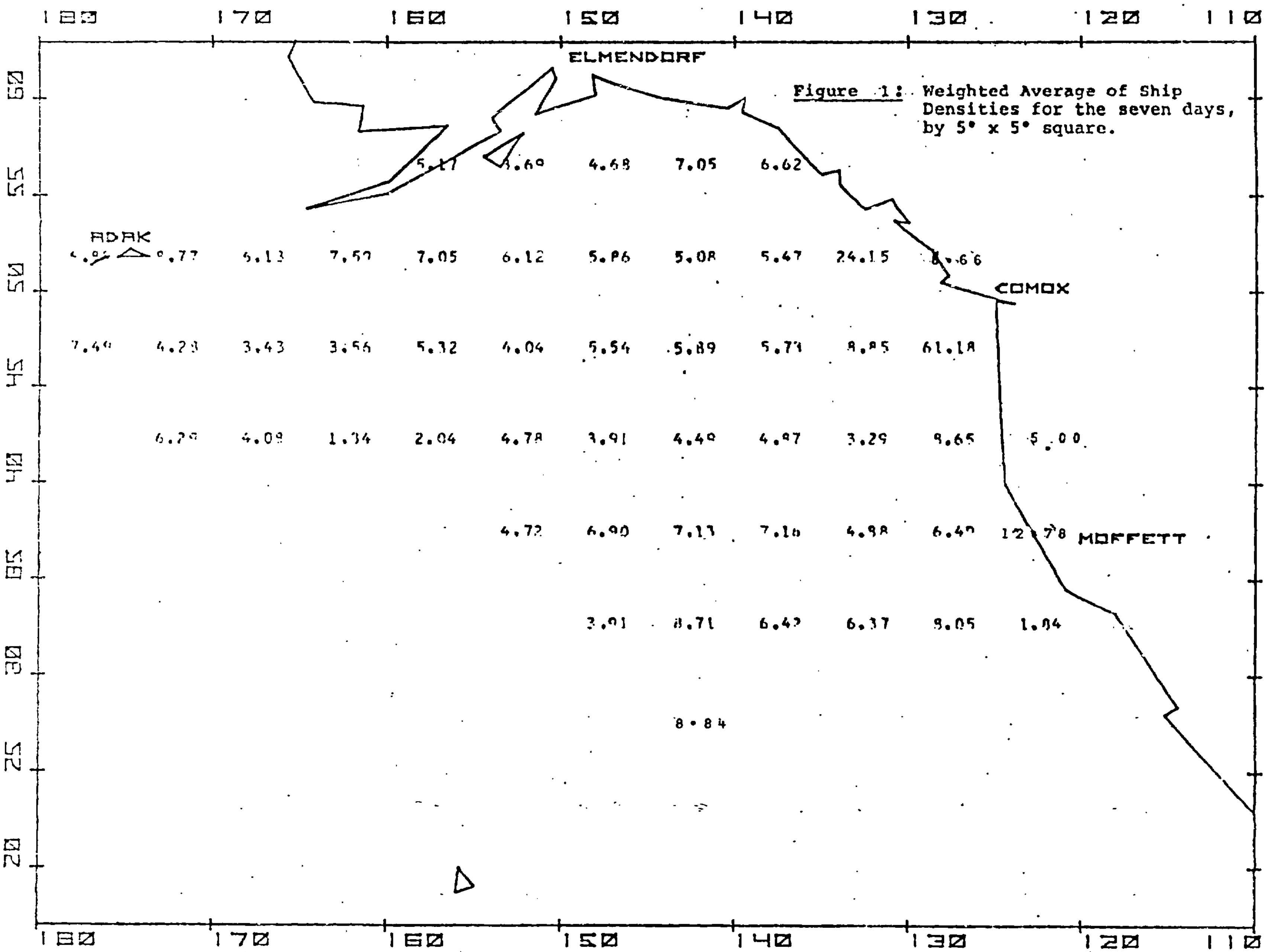


Figure 11: Weighted Average of Ship Densities for the seven days, by 5° x 5° square.

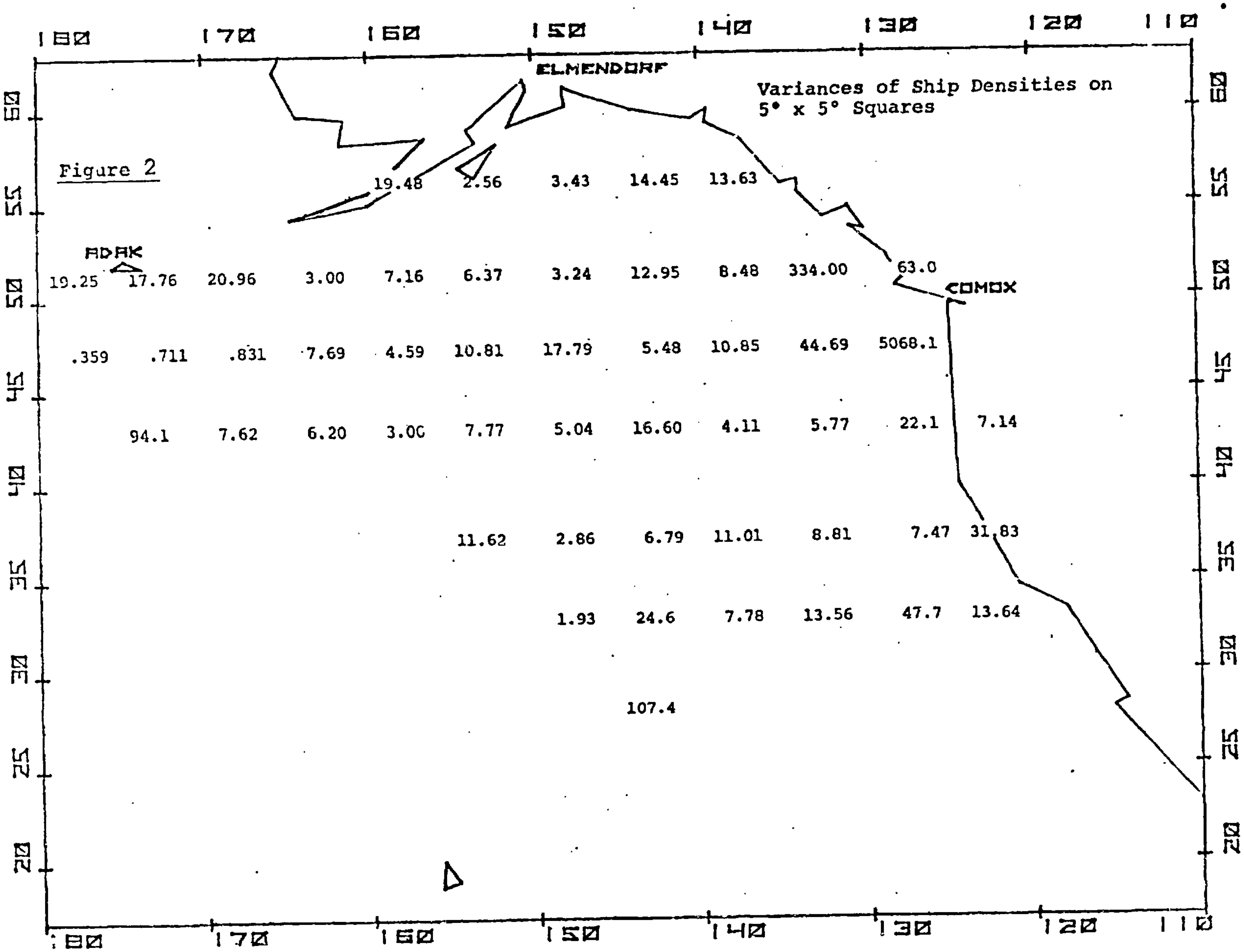


Figure 2

Variances of Ship Densities on 5° x 5° Squares

III. Algorithm for Estimating the Monthly Shipping Density Averages from a Partial Data Set

A. Introduction

The shipping density $D(x, y, t)$ at a given ocean locality is a stochastic function of time. Conceptually, it is defined as a number of ships per "unit" area. In practice, the unit area is taken to be a $1/2° \times 1/2°$ square in the ocean, thus $D(x, y, t)$ for any point in the grid is simply the number of ships in the grid. Thus instead of a continuous varying point (x, y) in the ocean, we have

$$D(x, y, t) = D(S_i, t) \qquad (1)$$

where $(x, y) \in S_i$

and S_i, $i = 1, 2, \ldots, N$ is a partition of the given ocean area into $1/2° \times 1/2°$ squares. Our problem is to estimate the time averages

$$D(i) \triangleq \overline{D(S_i, t)} = \frac{1}{T_1 - T_0} \int_{T_0}^{T_1} D(S_i, t)\, dt \qquad (2)$$

In the present problem we take $T_1 - T_0$ = one month

Now $D(S_i, t)$ is observed for only a few points in time t_j, $j = 1, \ldots, M$. In fact, $M \leq 4$. Furthermore the set of S_i observed at a given time t_j may or may not cover the entire ocean area. We shall assume that there is at least one time, t_1, for example, when the observation (by radar), covers the entire ocean area. The observation of the ocean area at a given time t_j may be represented by an $N \times M$ matrix C, i.e.,

$$C(i, j) = [C(S_i, t_j)] \quad i = 1, 2, \ldots, N;\ j = 1, \ldots, M \qquad (3)$$

where $C(S_i, t_j)$ = number of times square S_i is observed.

Of course $C(i, j) \geq 0$. And we stipulate that for $j = 1$, $C(i, j) \geq 1$. Corresponding to $C(i, j)$, we have the ship matrix S, i.e.,

$$S(i, j) = \text{number of ships reported to be in } S_i \text{ at time } t_j$$

Then the Density Matrix D is given by

$$D(i, j) = S(i, j)/ C(i, j) \tag{4}$$

where only elements of $D(i, j)$ for which $C(i, j) \neq 0$ are defined. Thus D is an incomplete matrix. The problem is then to estimate the vector $\vec{D}$,

$$\vec{D} = (D(1),\ldots D(N)) \text{ from the matrices C and S.}$$

B. Solution

The following approach is adopted.

Step 1: Complete the matrix D from the data C and S, by method of least square.

Step 2: $\hat{D}(i) = \frac{1}{M}[\sum_{j=1}^{M} w_j D(i, j)]$, (1)

where "^" denotes "estimation of" and the w_j's are weights.

C. Details of Step 1

Recall that the first column of D is complete, and for simplicity, we assume that all other columns of D are incomplete; we shall complete the other incomplete columns so as to maximize the correlation of the j-th column of D

$$\vec{D}_j = (D(i, j), \quad i = 1, 2,\ldots,N) \text{ with the first}$$

column, and with each other. And this is done by means of weighted least square. More precisely, we maximize the following function ρ.

$$\rho = \sum_{j=2}^{M} k_j \, \rho(D_1, D_j) + \sum_{\substack{i \geq 2 \\ i \neq j}}^{M} k_{ij} \, \rho(\vec{D}_i, \vec{D}_j) \tag{2}$$

where $\rho(\vec{D}_i, \vec{D}_j)$ is the proportionality factor between $\vec{D}_i$ and $\vec{D}_j$ and k's are some sort of weighting factors. The "non-existent" D(i, j) are variables to be estimated by maximizing ρ, i.e., by setting

$$\frac{\partial \rho}{\partial x} = 0 \tag{3}$$

One equation for each missing D(i, j) (= x) and k_j are some weighting factors. However to simplify the calculation, we tentatively set $k_{ij} = 0$.* Then ρ is maximized by separately maximizing

$$\rho_j = \rho(\vec{D}_1, \vec{D}_j) \tag{4}$$

i.e., by setting

$$\frac{\partial \rho_i}{\partial x} = 0 \tag{5}$$

one equation for each missing D(i, j) (= x) in $\vec{D}_j$

We further specify the form of ρ_j so that for a given missing D(i, j) (= x), equation

$\frac{\partial \rho_i}{\partial x} = 0$ is of the following form

$$\frac{\partial}{\partial x} \left[\sum_{i \varepsilon E_j} C(i, 1) \; C(i, j) \; (D(i, 1) \; x - D(i, j) \cdot x_1)^2 \right] \tag{6}$$

where E_j = the set of i ε[1,2,...,N] for which D(i, j) exist and x_1 = D(i, 1), the element in the first column corresponding to to the missing element in the $j^{\underline{th}}$ column

$$x = D(i, j).$$

*From the fact that many elements of $\vec{D}_i$, $\vec{D}_j$ are missing, we expect $k_{ij} << k_i, k_j$. Hence as a first approximation, $k_{ij} = 0$.

The complete form of ρ_j is

$$\rho_j = \sum_{\substack{x,y,z \\ \cdots}} \left(\sum_{i \varepsilon E_j} C(i, 1)\ C(i, j)\ D(i,1)\ x - D(i, j)\ x_1 \right)^2 \tag{7}$$

From (6) we deduce immediately*

$$\frac{x}{x_1} = \frac{\sum_{i \varepsilon E_j} C(i, 1)\ C(i, j)\ D(i, j)^2}{\sum_{i \varepsilon E_j} C(i, 1)\ C(i, j)\ D(i, 1)\ D(i, j)} = a_j \tag{8}$$

Thus we see that a_j is the same for all $x = D(i, j)$ in $\vec{D}_j$.

Thus we have simply

$$D(i, j) = a_j\ D(i, 1) \tag{9}$$

for all $i \not\varepsilon E_j$. Thus the missing $D(i, j)$ can be estimated for each $j = 2,\ldots,M$ and the monthly average computed by

$$D(i) = \left(\sum_{j=1}^{M} D(i, j)\right)/M \tag{10}$$

*This is the benefit of assuming $k_{ij} = 0$, otherwise we have to solve a simultaneous system of linear equations or use an iterative method.

D. General Algorithm

Section C treated the special case where it was assumed that the first column of the D matrix (which corresponded to flight day 'one') was complete, and that all other columns were incomplete. That is, it was assumed that one of the flight days obtained total coverage of the area and all other days only had partial coverage. This is the simplest case, and was used to illustrate the technique involved. In practice, of course, the situation is much more complex. The number of complete area coverage days may be much greater than one, or it may be zero.

Also, the simplified case presented above assumed that all days had equal weighting; this is rarely the case, for reasons mentioned in section I-B. Thus appropriate weights must be computed. As was done in section II, weights are determined by the number of times each flight covered the specific squares which it was surveying.

IV. Theory of Estimating Ship Density Time Averages for Incomplete Data

A. Introduction

The previous section presented an algorithm for estimating ship density time averages from incomplete data. It was a weighted least square method, minimizing a certain weighted sum of squares. It turns out that the computation of that algorithm is quite simple. In this paper we tackle the same problem from a theoretical point of view. This naturally leads to a maximal likelihood estimate for the density. It is shown that the maximal likelihood estimate will tend to make each component of the cited sum of squares small, thus vindicating our algorithm.

The exact numerical solution of the maximal likelihood estimate is much more complicated than the algorithm of Section III. The algorithmic solution is utilized as an approximation to the maximal likelihood estimate.

The notation of this section is consistent with that used in the description of the algorithm. In particular the meaning of C, D, $\vec{D}$, $\vec{D}_j$, S, S_i, M, N,... are the same.

B. Theory, Assumptions and Notation

Recall that $\bigcup_{i=1}^{N} S_i$ is a partition of the ocean area of interest. Observations are made at times t_j, j=1,...,N. The ship matrix S looks schematically as follows

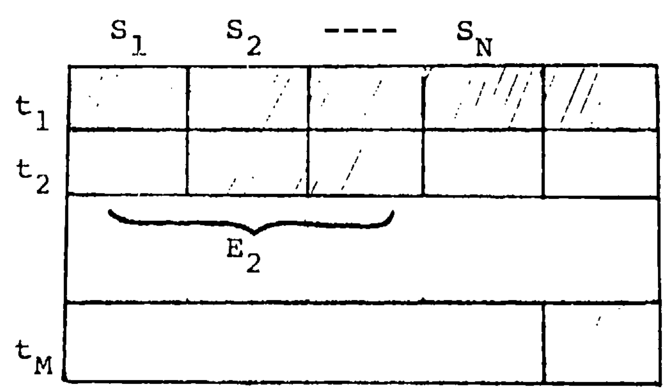


Fig. 3 Schematic View of S

where E_j is the set of S_i's for which D(i,j) exists.

We shall <u>assume</u> that for each square S_i, there is a probability $p(S_i) = p_i$ such that given a ship in the area, it will be found to be in the square S_i. We assume that p_i is constant in the period from t_1 to t_M (i.e., we assume that the relative densities throughout the area do not change in the time period under consideration.) By looking at the generating function

$$g(x_1,\ldots,x_N|K) = (p_1x_1 + \ldots + p_Nx_n)^K \tag{1}$$

where

$$\sum_{i=1}^{N} p_i = 1 \tag{2}$$

It is seen that the probability of there being k_i ships in S_i, $i=1,\ldots,N$ is

$$P_r(k_1,\ldots,k_N) = C(K;\ k_1,\ldots,k_N)\prod_{i=1}^{N}(P_i)^{k_i} \tag{3}$$

where

$$\Sigma\ k_i = K$$

and C is the multinomial coefficient.

Let there be a total of K ships in the area and assume that only a subset E of the squares in area S have been observed that day (either full or partial coverage of the square). For convenience let the subset E of

$$S = \{S_1,\ S_2,\ldots,S_N\}$$

be given by

$$E = \{S_i : i\ \varepsilon\ Q\}$$

where Q is a subset of $\{1,2,\ldots,N\}$.

Define

$$p_E = \prod_{i\varepsilon Q} p_i$$

$$K_E = \sum_{i \varepsilon Q} k_i$$

$$p_C = 1 - p_E$$

$$K_C = K - K_E$$

Then the probability that there are k_i ships in square S_i for all squares S_i in the subset E (gives a total of K ships) is

$$\Pr[D(S_i) = k_i \; \forall \, i \varepsilon Q | K]$$

$$= \binom{K}{K_E} (p_E)^{K_E} (p_C)^{K_C} \cdot \frac{(K_E)!}{\prod_{i \varepsilon Q} k_i!}$$

$$\cdot \prod_{i \varepsilon Q} (p_i/p_E)^{k_i}$$

$$= \frac{K!}{K_C! \prod_{i \varepsilon Q} k_i!} (p_C)^{K_C} \prod_{i \varepsilon Q} (p_i)^{k_i} \; . \qquad (4)$$

Denote this function by $\Pr(Q|K)$, which is a much simpler, though nonrigorous (and somewhat sloppy) notation.

C. Maximal Likelihood Estimates

The Likelihood function L for the M time observations in the overall area S is given by

$$L(S) = \log \prod_{j=1}^{M} \Pr(Q_j | K_j) \qquad (1)$$

where Q_j is the subset of indicies corresponding to the square observed on the j-th day, and K_j is the total number of ships in area S on the j-th day.

That is, if E_j denotes the subset of squares S_i of the set S which were covered on the j-th day of observation, then

$$E_j = \{S_i : i \varepsilon Q_j\}.$$

The unknowns in equation (1) are the total number of ships in the area for each day of coverage (note this is not equal to the number of ships observed on those days), and the base probabilities for each square.

These unknowns are determined as those values which maximize L(S), i.e., the solution of the following system

$$\left.\begin{aligned} &\frac{\partial}{\partial K_j}[L(S) + \lambda(1 - \sum_{i=1}^{N} p_i)] = 0 \\ &\qquad\qquad\qquad\qquad \text{for} \quad j \varepsilon J \\ &\frac{\partial}{\partial p_i}[L(S) + \lambda(1 - \sum_{i=1}^{N} p_i)] = 0 \\ &\qquad\qquad\qquad\qquad \text{for} \quad 1 \leq i \leq N \end{aligned}\right\} \quad (2)$$

where, of course, we have the constraining relationship

$$\sum_{i=1}^{N} p_i = 1$$

The set J consists of indicies of all days which did not cover (full or partially) every square, i.e.

$$J = \{j : 1 \leq j \leq M \text{ and } E_j \neq S\}.$$

Or, more explicitly, the system of equations is:

$$\frac{\partial}{\partial K_j}[\log K_j! - \log \hat{\kappa}_j! + \hat{\kappa}_j \log (1-q_j)] = 0$$
$$\text{for} \quad j \varepsilon J$$

$$\frac{\partial}{\partial p_i}\Big[\sum_{i=1}^{N}\ \sum_{j\varepsilon\Phi} D(i,j)\ \log\ p_i\Big]$$

$$+\ \frac{\partial}{\partial p_i}\Big[\sum_{j\varepsilon J}\ \kappa_j\ \log\ q_j\ +\ \sum_{j\varepsilon J}\ \hat{\kappa}_j\ \log\ (1-q_j)\Big]$$

$$+\ \sum_{j\varepsilon J}\ \frac{\partial}{\partial p_i}\ \Big[\sum_{i\varepsilon Q_j}\ D(i,j)\log\ p_i\Big]\ -\ \lambda\ =\ 0 \qquad (3)$$

$$\text{for } 1\leq i\leq N$$

where we define (using notation from section III)

$$K_j\ \equiv\ D(j)\ =\ \sum_{i=1}^{N}\ D(i,j)$$

$$\kappa_j\ \equiv\ D_E(j)\ =\ \sum_{i\varepsilon Q_j}\ D(i,j)$$

$$\hat{\kappa}_j\ =\ K_j\ -\ \kappa_j$$

$$\Phi\ =\ \{j:1\leq j\leq M\}\ -\ J$$

$$q_j\ =\ \sum_{i\varepsilon Q_j}\ p_i$$

That is, κ_j is the partial sum of ships, $\hat{\kappa}_j$ its difference from the total sum K_j, Φ is the complement of J and q_j is the partial sum of probabilities. The quantity λ is the Lagrangian multiplier arising from the restraining equation.

Without loss of generality we may assume that the set J has exactly one element, and this element will be taken to be '1'. For it is assumed that on some day all squares will be covered (fully or partially); if this is not so, historical data can be utilized to establish a 'base' field, or first overall estimate. (Incorporation of historical data bases with observational data

in these algorithms can be easily accomplished. However under such conditions weights should definitely be used to reflect the relative confidence in the differing data types - see section I). If J consists of more than one element, then those days may be combined (using appropriate weighting, as is done in section II), into a single 'day' of data, which is then utilized. In practice, of course, the algorithm given in section III performs properly regardless of the size of this artificial set 'J'. Therefore, there is no loss of generality in assuming that the set J consists of the single index '1', which is implicitly used in the following subsections.

We shall solve the problem first for the case M = 2, then for the case M = 3. These two specific cases will suffice for illustrative purposes; in practice the least squares algorithm is always used, which simplifies computation immensely.

D. <u>The Case M = 2</u>

To simplify notation we write q, κ, K and E for q_2, κ_2, K_2 and E_2 respectively. The requirement is to solve the following simultaneous system:

$$\frac{\partial}{\partial K}[\log K! - \log(K-\kappa)! + (K-\kappa)\log(1-q)] = 0$$

$$D(i, 1)/p_i + [\frac{\kappa}{q} - \frac{K-\kappa}{1-q} + \frac{D(i,2)}{p_i}]\ \phi_E(i) = \lambda \qquad \text{for } 1 \le i \le N$$

$$\sum_{i=1}^{N} p_i = 1 \tag{4}$$

where ϕ is the characteristic function of the set E, i.e.

$$\phi_E(i) = \begin{cases} 1 \text{ if } i \in E \\ 0 \text{ if } i \notin E \end{cases}$$

The first equation will be replaced by an approximation by using the Stirling approximation for K!

$$K! \approx \sqrt{2\pi K}\, K^K \exp(-K + \omega(K)), \qquad (5)$$

where

$$\frac{1}{12K+6} < \omega(K) < \frac{1}{12K}$$

Hence

$$\frac{\partial}{\partial K} \log \frac{K!}{(K-\kappa)!} \approx \frac{1}{12} \frac{\kappa(2K-\kappa)}{(K-\kappa)^2\, K^2} - \frac{1}{2} \frac{\kappa}{K(K-\kappa)}$$

$$+ \log \frac{K}{K-\kappa}$$

Thus K is a root of the system:

$$\frac{1}{12} \frac{\kappa(2K-\kappa)}{K^2 (K-\kappa)^2} - \frac{1}{2} \frac{\kappa}{K(K-\kappa)} + \log \frac{K}{K-\kappa} + \log(1-q) = 0$$

$$D(i,1) = \lambda p_i \quad \text{for } i \notin Q$$

$$D(i,1) + \left[\left(\frac{\kappa}{q}\right) - \left(\frac{K-\kappa}{1-q}\right) + \frac{D(i,2)}{p_i}\right] p_i = \lambda p_i$$

$$\text{for } i \in Q \qquad (6)$$

The system (6) can be solved iteratively as follows.

Adding the last equations in the system for $i=1,\ldots,N$ we get

$$D(1) + D_E(2) + \kappa - \frac{K-\kappa}{1-q} \cdot q = \lambda \qquad (7)$$

As an initial approximation to q, we take

$$(q)^1 = D_E(1)/D(1) \qquad (8)$$

Substituting this value of q into the first equation of (6) we solve for K and denote the estimate by $(K)^1$. Substituting this value of K into eq. (7), we get $(\lambda)^1$, hence we get

$$(p_i)^1 = D(i,1)/(\lambda)^1 \quad \text{for } i \notin Q$$

and

$$(p_i)^1 = [D(i,1) + D(i,2)]/((\lambda)^1 - (F)^1) \qquad (9)$$

where

$$(F)^1 = -\frac{(K)^1 - \kappa}{1 - (q)^1} \cdot (q)^1$$

Then

$$(q)^2 = \sum_{i \in Q} (p_i)^1$$

And the iteration continues. Under suitable conditions the iteration will converge, giving the final estimates of

$$\hat{q} = \lim_{n\to\infty} (q)^n$$

$$\hat{\lambda} = \lim_{n\to\infty} (\lambda)^n$$

$$\hat{p}_i = \lim_{n\to\infty} (p_i)^n \qquad (10)$$

Of course, as with most iterative procedures designed to solve maximum likelihood estimation problems, it is possible for the data to be such that the estimates either diverge, oscillate, or converge to a false position which is not the maximum.

Since we are only interested in the theoretical properties of the problem, and do not intend to use this procedure for actual computation in a production environment, it is of no value to digress into techniques of numerical analysis to handle such cases, or to examine the behavior of the system when the maximum lies on the boundary of the domain, rather than the interior case (which is the general case). It suffices to say that the M.L.E. function is continuous on a compact subset of Euclidean space, therefore a maximum exists.

E. The Relation of Maximal Likelihood Estimates to the Least Square Method

Recall that the least square estimates $\widehat{D(i,2)}$, $i \notin Q$

are obtained by minimizing each of the following expressions

$$\rho_\ell = \sum_{i \in Q} C(i,1) \cdot C(i,2) [D(\ell,1) D(i,2) - D(\ell,2) D(i,1)]^2 \quad \text{for } \ell \notin Q \tag{1}$$

where the C's and D's are elements of the matrices used in section XII.

To digress a minute, consider the random variable

$$Z = \sum_{j=1}^{K} \xi_j \tag{2}$$

where the ξ's are independent Bernoulli random variables with parameter (expectation) p_ℓ and K is an integer valued random variable. Then this is a stochastic representation of the number of ships in the ℓ-th square when the variable K denotes the total number of ships in the area. Now the minimizing value of $D(\ell,2)$ for $\ell \notin Q$ (see equation 1 above) is the estimator of the expected value of Z, which is

$$E[Z] = p_\ell E[K] \tag{3}$$

[In the situation where there is partial coverage of squares (or multiple coverage of part of a square by different flights), then the above heuristic model must be expanded, since now the estimator K need not be an integer. In such a case the modification is

$$Z = \sum_{j=1}^{k} \xi_j + \eta$$

where k is the integral portion of K and η is a discrete random variable related to the fractional portion of the random variable K. The expected value relation (3) still holds however.]

Now the maximal likelihood estimates of K and the p_ℓ's are such as to make Kp_ℓ approximate $D(\ell,2)$

and $\qquad D(1)p_\ell$ approximate $D(\ell,1)$, i.e.,

$$\left.\begin{array}{l} D(i,2) \approx Kp_i \\ D(i,1) \approx D(1)p_i \end{array}\right\} \text{ for } i \varepsilon Q$$

$$\left.\begin{array}{l} D(\ell,1) \approx D(1)p_\ell \\ D(\ell,2) \approx Kp_\ell \end{array}\right\} \text{ for } \ell \notin Q \tag{4}$$

Thus system (4) will make the term

$$[D(\ell,1)D(i,2) - D(\ell,2)D(i,1)]^2$$

small for all $i\varepsilon Q$ and $\ell \notin Q$, and hence it will make ρ small.

This shows that the least square estimates of $D(\ell,2)$, $\ell \notin Q$ must approximate the maximal likelihood estimates. However the numerical calculation of $D(\ell,2)$ by weighted least squares is much shorter than that of maximal likelihood estimates.

F. <u>The Case N = 3</u>

We shall illustrate the calculation of maximal likelihood estimates for $K_2,\ldots,K_M,p_1,\ldots p_N$ for the case $M = 3$. The calculations are much more complicated than the case $M = 2$. Thus we appreciate even more the quick method of least squares.

Corresponding to eq. (4) of subsection D, we have

$$\frac{\kappa_j(2K_j-\kappa_j)}{12K_j^2(K_j-\kappa_j)^2} - \frac{K_j}{2K_j(K_j-\kappa_j)} + \log\frac{K_j}{K_j-\kappa_j} + \log(1-q_j) = 0$$

$$j = 2,3 \tag{1}$$

For $i \notin Q_2 \bigcup Q_3$, we have

$$D(i,1) = \lambda p_i$$

For $i \in Q_2 - Q_3$, we have

$$D(i,1) + D(i,2) + [\frac{\kappa_2}{q_2} - \frac{K_2 - \kappa_2}{1 - q_2}] p_i = \lambda p_i \tag{2}$$

For $i \in Q_3 - Q_2$, we have

$$D(i,1) + D(i,3) + [\frac{\kappa_3}{q_3} - \frac{K_3 - \kappa_3}{1 - q_3}] p_i = \lambda p_i \tag{3}$$

For $i \in Q_2 \bigcap Q_3$, we have

$$D(i,1) + D(i,2) + D(i,3) + [F_2 + F_3] p_i = \lambda p_i \tag{4}$$

where $$F_2 = \frac{\kappa_2}{q_2} - \frac{K_2 - \kappa_2}{1 - q_2}$$

and a similar expression for F_3.

If we add equations (1) for i=1,...,N, we get

$$\lambda = D(1) + D_{E_2}(2) + D_{E_3}(3) + (F_2)q_2 + (F_3)q_3$$

We initialize q_2, q_3 by setting

$$(q_j)_1 = D_{E_j}(1)/D(1)$$

Then from equation (1) we solve for K_2, K_3 to get $(K_2)_1$ and $(K_3)_1$. Then equations (2), (3), (4) will get us $(p_i)_1$ from which $(q_j)_2$ are computed and the process repeated until sufficient accuracy is achieved.

V. Analysis of a Specific Case

This partial data set algorithm was utilized in the analysis of data obtained in the Alboran basin. The flight tracks varied widely, resulting in partial coverage for almost all flight days; in some instances, less than a quarter of the squares were covered (although this was very unusual).

There were three types of flights that gathered data on the Alboran during the exercise:

(a) Aircraft Surveillance flights performed during August.
(b) Aircraft Surveillance flights performed during November.
(c) Other flights which performed shipping surveillance. These flights had as their primary tasks other duties which allowed shipping data to be gathered at the same time. These covered a period of several months.

There were a few flights in groups (a) and (b) and several in group (c). Quality of the data varied widely; most flights in groups (a) and (b) obtained excellent data, while several flights in group (c) had radar difficulties. The time period of interest was November; thus, August shipping densities were scaled to reflect the lighter traffic in that region during November. Also they were weighted much lighter than the November flights. The representative density obtained for November is shown in figure 3. The confidence in the results depends upon the underlying data. In the Alboran Sea itself, much data was available and the results are felt to be very trustworthy. In the extreme eastern region of the area (squares 12-14, 22-24, 33-35, 46-47) little data was available

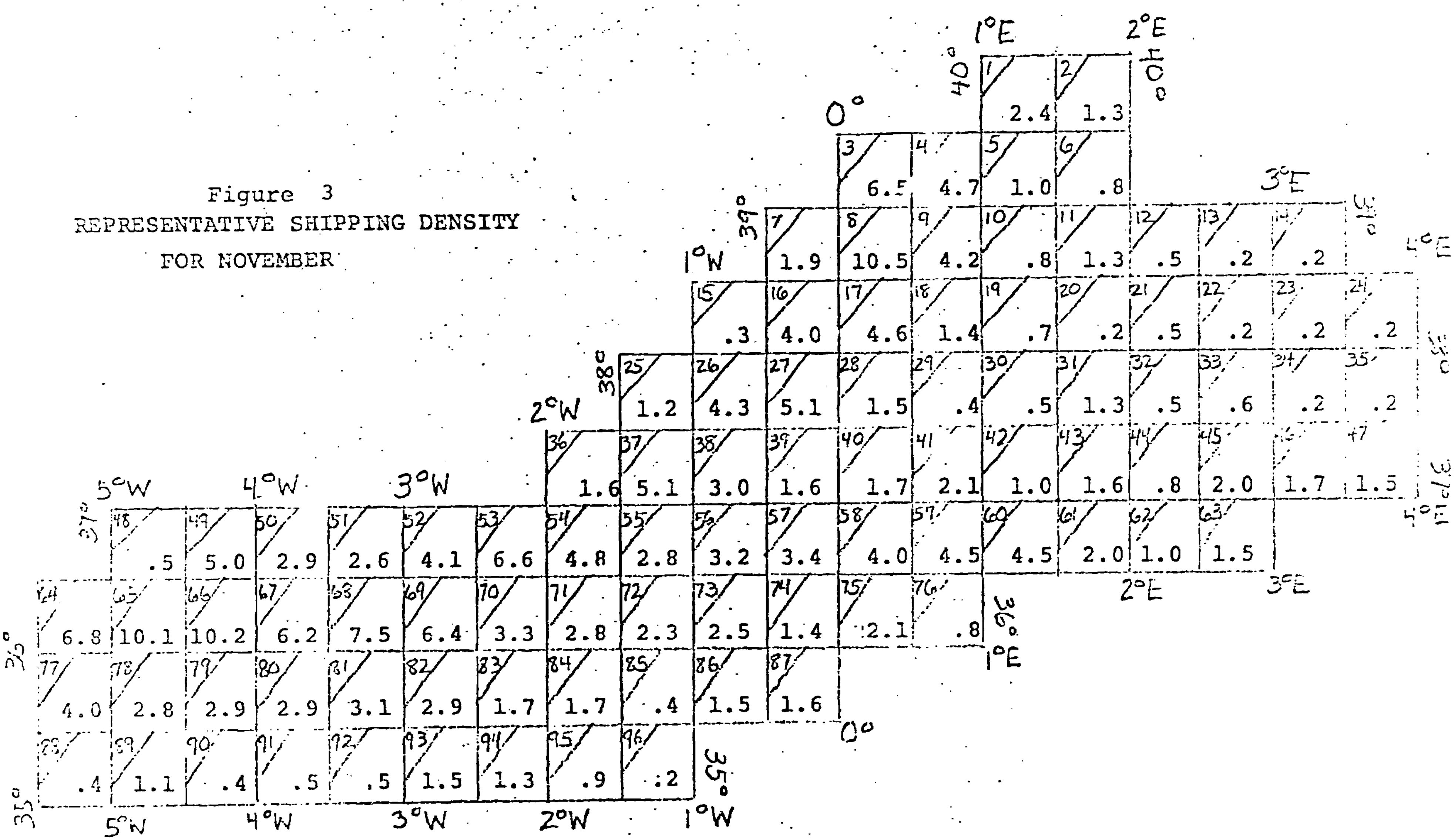
(only two flights gathered data on these squares); thus, historical data for these squares was entered as well as the observed data. Confidence for these end squares naturally is much lower than for the important areas.

A fishing fleet was observed on one flight of group (c), which naturally produced a very high density for that square on one day. Since the algorithm maximized overall correlation, the resultant shipping field did not show this abnormality in the square.

It is interesting to note that densities were computed both with and without the final data set (taken 2 December), and they agreed within 3%, indicating that the additional data was (statistically) very similar to the other data sets.

This algorithm has provided realistic shipping density estimates based on actual exercise data of vastly varying quality. It has shown relative insensitivity to singularly high data points which are inconsistent with the remainder of the data. Therefore, we feel that this algorithm will give an accurate and robust estimate of the shipping field under typical conditions encountered in actual exercises.

Figure 3
REPRESENTATIVE SHIPPING DENSITY
FOR NOVEMBER



VI. Theoretical Sensitivity Analysis

In this section we consider the sensitivity of the weighted average algorithm. To do this a theoretical stochastic model will be utilized, and the statistics of the estimator (i.e., the computed density) will be examined.

If an area of open ocean is small (in comparison to the world), then the number of ships in that area at a fixed time can be modelled by a Poisson distribution whose parameter λ is the expected number of ships in the area. As before, this small area is taken to be a square of suitable dimensions. The aircraft surveillance days are sufficiently separated that a ship in the square on one day will have left the square by the next surveillance day, thus we may assume that the ship counts on different days are independent (the basic premise of statistical sampling).

Therefore consider a square which contains (on the average) λ ships, and let these be η surveillance flights. Let p_i be the coverage of the i-th flight. Let X_i be a sequence of independent random variables having Poisson distributions with parameters λp_i respectively. Then the weighted average (section II, equation 6) is represented by the random variable

$$D = \Sigma X_i / s \qquad (1)$$

where

$$s = \Sigma p_i$$

If Z is a random variable having a Poisson distribution with parameter λs, then D is a law equivalent to Z/s, i.e., they have the same distribution

$$D \overset{L}{=} Z/s \qquad (2)$$

Since we are only interested in the distribution, the cumbersome term 'law equivalent' will be replaced by the common 'equals'.

The random variable $\mathcal{Z}$ is Poisson, hence

$$E(D) = \lambda \tag{3}$$
$$Var(D) = \lambda/s$$

so that as s tends to infinity, D tends to the constant λ almost surely,

$$\lim_{s\to\infty} D = \lambda \tag{4}$$

Consider the random variable

$$Y = \sqrt{s}\ (D - \lambda)$$

where characteristic function is

$$\phi_Y(t) \equiv E(e^{iYt}) = \exp\{\lambda s(e^{it/\sqrt{s}} - 1) - \lambda it\sqrt{s}\}$$

Now

$$\lim_{s\to\infty} \phi_Y(t) = \exp - \lambda t^2/2$$

uniformly (in t) on compact sets, hence the random variable Y converges in law to a normally distributed random variable with mean 0 and variance λ.

Therefore for large values of s, the distribution of D may be approximated by a normal distribution with mean λ and variance λ/s (compare with equation 3).

To evaluate the algorithm, a natural question to ask is: Given that the density is λ, where do the .025 and .975 points on the distribution of D lie? That is, find the largest d_L and the smallest d_U such that

$$\begin{aligned} Pr(D < d_L) &\leq .025 \\ Pr(D > d_U) &\geq .975 \end{aligned} \tag{5}$$

Then we have

$$Pr(D \epsilon [d_L, d_U]) \geq .95$$

i.e., 95% of the time the algorithm would give a value in this interval For large values of s we may use the normal approximation to obtain the value

$$d_L = \lambda - 1.96\sqrt{\lambda/s}$$

$$d_U = \lambda - 1.96\sqrt{\lambda/s}$$

For small values of s these limits should be computed directly from the Poisson distribution. If we only consider the relative bounds (i.e., by dividing by the mean), then we can express the equations in terms of the single parameter β where

$$\beta = \lambda s$$

viz.,

$$\delta_L = d_L/\lambda = 1 - 1.96/\sqrt{\beta}$$
$$\delta_U = d_U/\lambda = 1 + 1.96/\sqrt{\beta} \tag{6}$$

Figure 4 graphs these two normally approximated bounds, as well as giving the actual bounds from the Poisson distributed variable Z. Note that the Poisson (relative) bound, being an integral multiple of $1/\beta$, does not converge monotonically as the normal bound does, but is discontinuous.

Now consider a problem which is quite different yet often confused with the previous one. Specifically, given an estimate $\hat{\lambda}$ of the density in a square (with a fixed coverage s), construct a confidence interval for the actual density λ. That is, we wish to find the greatest lower bound λ_L such that if this (or any smaller value) is the true density, then the probability that the estimated density is $\hat{\lambda}$ or greater is less than a prespecified amount (e.g., .025). A similar statement holds for an upper

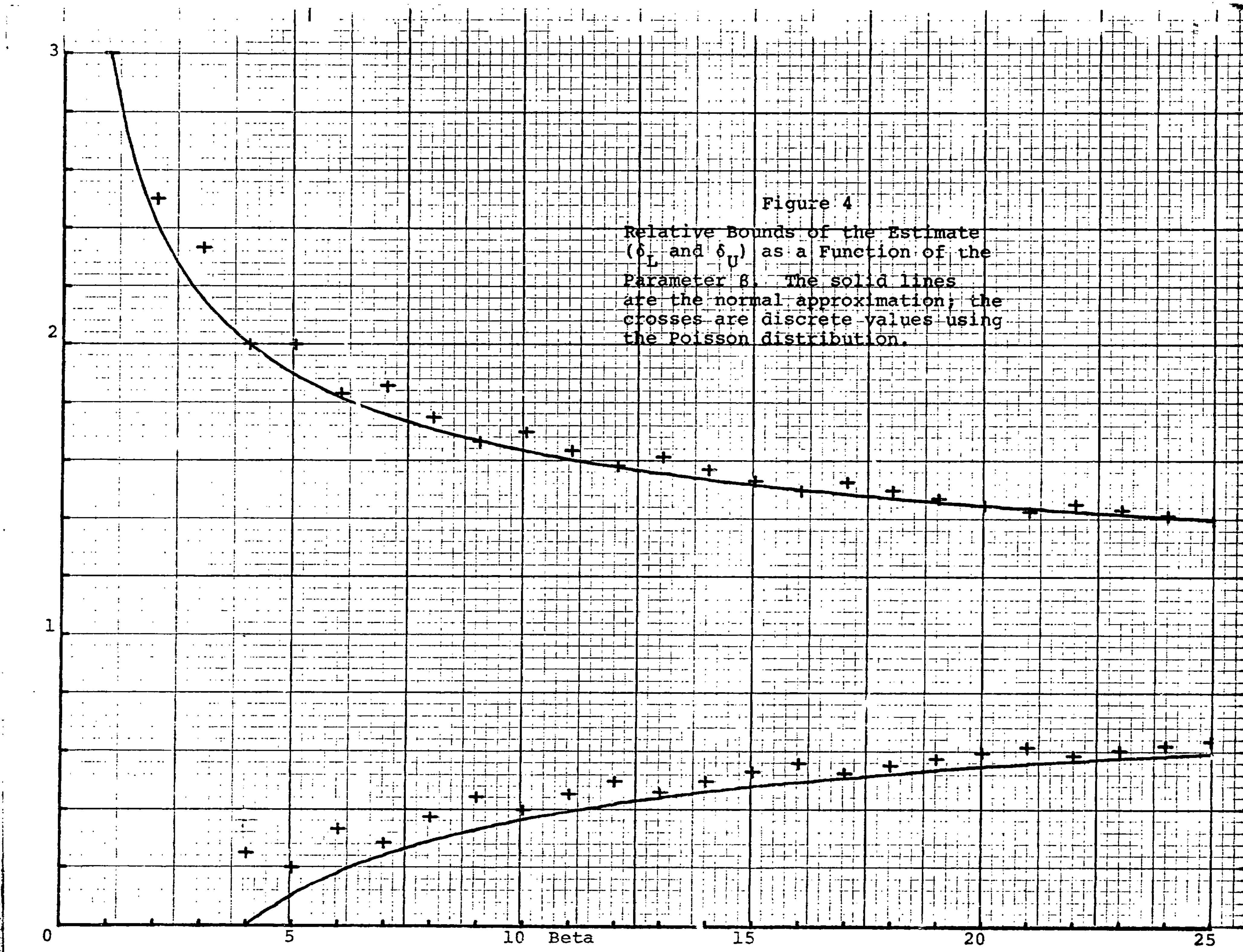


Figure 4

Relative Bounds of the Estimate (δ_L and δ_U) as a Function of the Parameter β. The solid lines are the normal approximation; the crosses are discrete values using the Poisson distribution.

bound λ_U. This can be expressed as the largest λ_L and the smallest λ_U such that

$$\begin{aligned} \Pr(D \geq \hat{\lambda} \mid \lambda \leq \lambda_L) &\leq .025 \\ \Pr(D \leq \hat{\lambda} \mid \lambda \geq \lambda_U) &\leq .025 \end{aligned} \tag{7}$$

Note that λ_L and λ_U are functions of both $\hat{\lambda}$ and s. Following the previous analysis we introduce new parameters, viz.,

$$\begin{aligned} \beta_L &= \lambda_L s \\ \beta_U &= \lambda_U s \\ \hat{\beta} &= \hat{\lambda} s \end{aligned} \tag{8}$$

Once again it turns out that there is only one essential parameter, for β_L and β_U may be expressed solely in terms of $\hat{\beta}$. If we recall the curves of Figure 4, where the abscissa was the parameter β, then the values of β_L and β_U (for a given $\hat{\beta}$) may be conceived graphically as the intersections with a rectangular hyperbola of parameter $\hat{\beta}$, as illustrated in Figure 5.

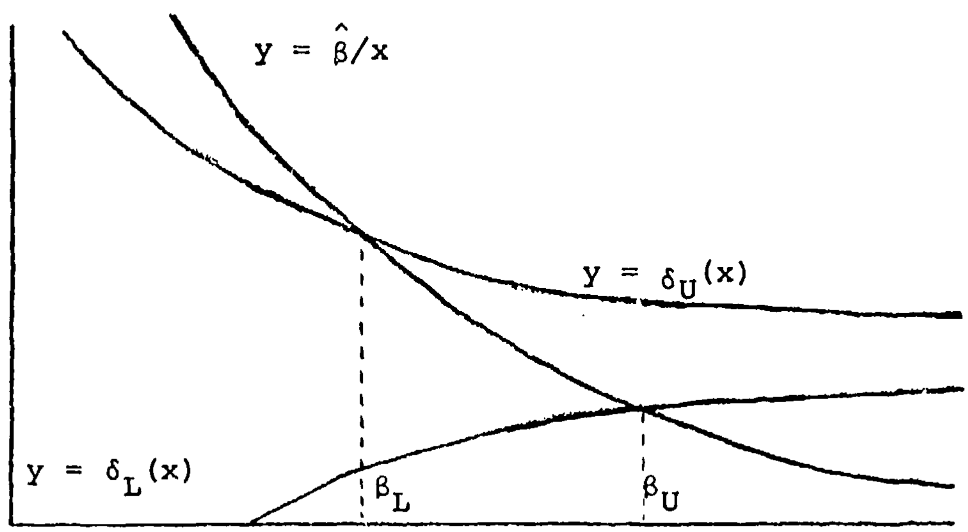


Figure 5
Geometrical Relationship of β_L and β_U upon $\hat{\beta}$

Analytically, of course, β_L is the solution of the equation

$$\hat{\beta}/\beta_L = \delta_U(\beta_L), \quad \beta_L \in (0, \infty)$$

and β_U satisfies the equation

$$\hat{\beta}/\beta_U = \delta_L(\beta_U), \quad \beta_U \in (0, \infty)$$

Solving these equations in the asymptotic case (i.e., using the normal distribution rather than the Poisson) yields

$$\beta_L = \{\sqrt{\hat{\beta} + .9604} - .98\}^2$$

$$\beta_U = \{\sqrt{\hat{\beta} + .9604} + .98\}^2 \qquad (9)$$

where, of course, all square roots are positive. Graphs of these functions are presented in Figure 6. Note that β_L and β_U are the two sections of the parabola

$$(\hat{\beta} - \beta)^2 = 3.8416\beta$$

From these graphs one may immediately obtain a confidence interval for the true density λ using the equations of system (8).

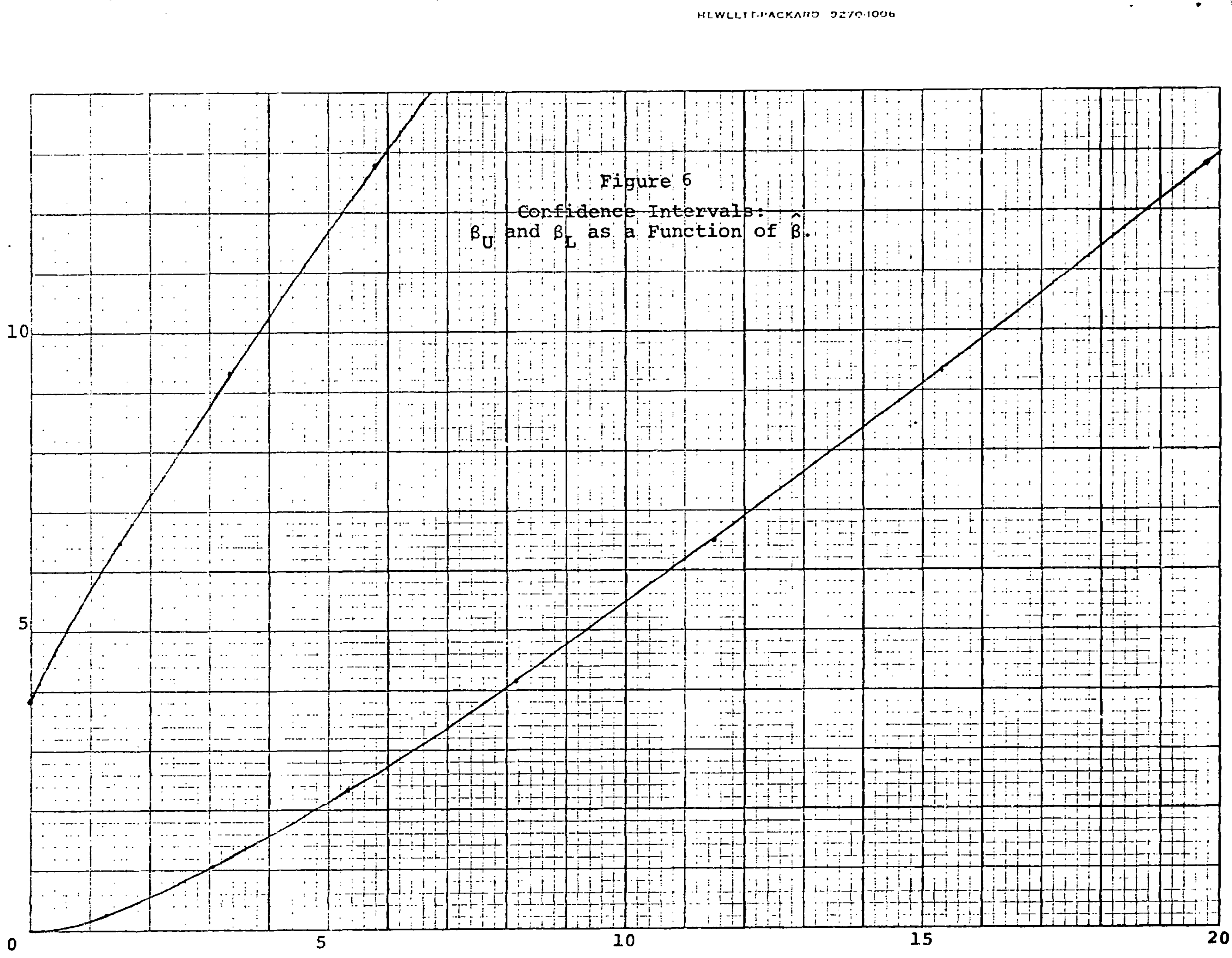


Figure 6

Confidence Intervals:
β_U and β_L as a Function of $\hat{\beta}$.

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
875 NORTH RANDOLPH STREET
SUITE 1425
ARLINGTON VA 22203-1995

IN REPLY REFER TO:

5510/1
Ser 321OA/011/06
31 Jan 06

MEMORANDUM FOR DISTRIBUTION LIST

Subj: DECLASSIFICATION OF LONG RANGE ACOUSTIC PROPAGATION PROJECT (LRAPP) DOCUMENTS

Ref: (a) SECNAVINST 5510.36

Encl: (1) List of DECLASSIFIED LRAPP Documents

1. In accordance with reference (a), a declassification review has been conducted on a number of classified LRAPP documents.

2. The LRAPP documents listed in enclosure (1) have been downgraded to UNCLASSIFIED and have been approved for public release. These documents should be remarked as follows:

   Classification changed to UNCLASSIFIED by authority of the Chief of Naval Operations (N772) letter N772A/6U875630, 20 January 2006.

   DISTRIBUTION STATEMENT A: Approved for Public Release; Distribution is unlimited.

3. Questions may be directed to the undersigned on (703) 696-4619, DSN 426-4619.

BRIAN LINK
By direction

Subj: DECLASSIFICATION OF LONG RANGE ACOUSTIC PROPAGATION PROJECT (LRAPP) DOCUMENTS

DISTRIBUTION LIST:

NAVOCEANO (Code N121LC – Jaime Ratliff)
NRL Washington (Code 5596.3 – Mary Templeman)
PEO LMW Det San Diego (PMS 181)
DTIC-OCQ (Larry Downing)
ARL, U of Texas
Blue Sea Corporation (Dr.Roy Gaul)
ONR 32B (CAPT Paul Stewart)
ONR 321OA (Dr. Ellen Livingston)
APL, U of Washington
APL, Johns Hopkins University
ARL, Penn State University
MPL of Scripps Institution of Oceanography
WHOI
NAVSEA
NAVAIR
NUWC
SAIC

Declassified LRAPP Documents

Report Number	Personal Author	Title	Publication Source (Originator)	Pub. Date	Current Availability	Class.
55	Weinstein, M. S., et al.	SUS QUALITY ASSESSMENT, SQUARE DEAL	Undersea Systems, Inc.	750207	ADA007559; ND	U
BKD2380	Unavailable	WESTLANT 74 PHASE 1 DATA SUMMARY (U)	B-K Dyanmics, Inc.	750301	NS; ND	U
TM-SA23-C44-75	Wilcox, J. D.	MOTION MODEL VALIDATION FROM LRAPP ATLANTIC TEST BED DATA	Naval Underwater Systems Center	750317	ND	U
RAFF7412; 74-482	Scheu, J. E.	SQUARE DEAL SHIPPING DENSITIES (U)	Raff Associates, Inc.	750401	ADC003198; NS; ND	U
PSI TR-004018	Barnes, A. E., et al.	ON THE ESTIMATION OF SHIPPING DENSITIES FROM OBSERVED DATA	Planning Systems Inc.	750401	ND AD 096 582	U
NUSC TD No.4937	LaPlante, R. F., et al.	THE MOORED ACOUSTIC BUOY SYSTEM (MABS)	Naval Underwater Systems Center	750404	ADB003783; ND	U
USI 460-1-75	Weinstein, M. S., et al.	SUS SIGNAL DATA PROCESSING (U) INVESTIGATIONS CONDUCTED UNDER THE DIAGNOSTIC PLAN FOR CHURCH ANCHOR AND SQUARE DEAL SHOT DATA (U)	Underwater Systems, Inc.	750414	ADC002353; ND	U
Unavailable	Ellis, G. E.	SUMMARY OF ENVIRONMENTAL ACOUSTIC DATA PROCESSING	University of Texas, Applied Research Laboratories	750618	ADA011836	U
Unavailable	Edelblute, D. J.	OCEANOGRAPHIC MEASUREMENT SYSTEM TEST AT SANTA CRUZ ACOUSTIC RANGE FACILITY (SCARF)	Lockheed Missiles and Space Co., Inc.	751015	ADB007190	U
Unavailable	Unavailable	SUS SOURCE LEVEL COMMITTEE REPORT	Underwater Systems, Inc.	751105	ADA019469	U
Unavailable	Hampton, L. D.	ACOUSTIC BOTTOM INTERACTION EXPERIMENT DESCRIPTION	University of Texas, Applied Research Laboratories	760102	ADA021330	U
PSI-TR-036030	Turk, L. A., et al.	CHURCH ANCHOR: AREA ASSESSMENT FOR TOWED ARRAYS (U)	Planning Systems Inc.	760301	ND	U
NUC TP 419	Wagstaff, R. A., et al.	HORIZONTAL DIRECTIONALITY OF AMBIENT SEA NOISE IN THE NORTH PACIFIC OCEAN (U)	Naval Undersea Center	760501	ADC007023; NS; ND	U
NRL-MR-3316	Young, A. M., et al.	AN ACOUSTIC MONITORING SYSTEMS FOR THE VIBROSEIS LOW-FREQUENCY UNDERWATER ACOUSTIC SOURCE	Naval Research Laboratory	760601	ADA028239; ND	U
ARL-TR-75-32	Ellis, G. E.	SUMMARY OF ENVIRONMENTAL ACOUSTIC DATA PROCESSING	University of Texas, Applied Research Laboratories	760705	ADA028084; ND	U
Unavailable	Unavailable	SUMMARY OF ENVIRONMENTAL ACOUSTIC DATA PROCESSING	University of Texas, Computer Science Division	761013	ADA032562	U
TTA83676285	Unavailable	ANALYSIS PLAN FOR NARROWBAND/ NARROWBEAM AMBIENT NOISE (U)	Tetra Tech, Inc.	761112	ADC008275; NS; ND	U
USI 564-1-77	Wallace, W. E., et al.	REPORT OF CW WORKSHOP. NORDA, BAY ST. LOUIS, MISS., 28-29 SEPT 1976	Underwater Systems, Inc.	770124	ND	U